Beyond the Average

The Evolutionary Importance of Gene Interactions and Variability of Epistatic Effects

PATRICK C. PHILLIPS
SARAH P. OTTO
MICHAEL C. WHITLOCK

The importance of interactions between genetic loci in evolutionary processes was a focus of disagreement between Fisher and Wright (Provine 1971, 1986). In particular, Fisher believed that the interactions of alleles at different loci are relatively unimportant and that the evolutionary future of an allele could be well summarized by its effect averaged across genetic backgrounds. Consequently, Fisher believed that the force of natural selection, rather than idiosyncratic interactions among genes, is paramount in determining the evolutionary trajectory of a population. Wright, on the other hand, believed that these epistatic interactions are central, as they define the valleys and peaks of the fitness landscape within which organisms evolve. Most famously, gene interactions played a critical role in Wright’s shifting-balance theory, which maintains that random drift and selection determine the fate of a population on hilly, epistatic fitness surfaces. This disagreement between these two scientists took many forms, but centered on the question of whether we must consider the details of how an allele interacts with other alleles or whether we can simply average all the details away (see also Goodnight, chap. 8, this volume; Templeton, chap. 3, this volume; Wolf, chap. 10, this volume).

Beyond the Fisher/Wright debate on the nature of adaptation, gene interactions are known to have a considerable influence on a number of evolutionary processes, including the evolution of sex, the evolution of recombination, mutation load, inbreeding depression, mutational meltdown, Muller’s ratchet, and the evolution of reproductive isolation among species. Each of these processes depends not only on the rate of mutation to new forms but also on the various types of interactions that these new mutants face. In almost every case where these evolutionary processes have been examined, however, the epistatic effects have been modeled as an average, with all genes interacting in a constant and uniform fashion.

Gene interactions, of course, do not occur evenly among all possible pairs of alleles. Certain combinations of alleles at different loci interact to cause phenotypes or fitnesses that are extremely different from those predicted by looking at the effects of each allele in isolation, while other allelic combinations do not. Given the nature of biological processes, it would indeed be surprising if such variation was not found. Variability in extent and effect of interactions among different combinations of loci has many important ramifications, but, most important, this variation indicates that we know little about the nature of a particular interaction from the mean interaction effect. Many of the important biological consequences of gene interactions depend on specific locus-to-locus interactions, rather than on some average level of interaction. Thus, we as a field may have significantly underestimated the importance of gene interactions by using average estimates of epistatic effects. Variance in epistatic effects has only recently received theoretical (e.g., Otto and Feldman 1997) and empirical (e.g., Elena and Lenski 1997) attention. Our primary goal here is to focus attention on this issue by illustrating the impact of variance in epistatic effects on evolutionary processes and by highlighting ways in which these effects can be experimentally investigated.

In this chapter, we begin to conceptualize gene interactions in a way that encompasses the complex interplay between loci that must exist in real systems. Not only do we look beyond Fisher’s view of the average effects of single loci, but also we try to look beyond the modern view of average levels of epistatic effects. Evolution occurs in a multidimensional genotypic space that cannot be justifiably reduced to a one-dimensional or two-dimensional representation. Wright, himself, understood the idiosyncratic nature of gene interactions and tried to capture these multidimensional interactions in several diagrams (cf. Wright 1977, chap. 13). This understanding came primarily from Wright’s experience with coat-color mutants and the specific ways in which these alleles interact (Provine 1986). Similarly, as we learn more about the context in which a gene acts, and about cell structure, metabolic pathways, chemical signaling, and developmental programs, it is becoming more and more obvious that evolutionary theory cannot ignore what genes do and the genetic context in which they do it.

We begin by reviewing concepts and data on gene interactions and then discuss how the variable nature of epistatic interactions may affect evolutionary processes, including mutation load, evolution of sex and recombination, drift load, and speciation.

Definitions

The term “epistasis,” as a description of the masking of the expression of one-locus by the alleles at another locus, was introduced into genetics by William Bateson towards the beginning of the 20th century (Phillips 1998). Fisher (1918) expanded...
on this concept to include quantitative differences among genotypes, calling any deviation from the additive combination of single-locus genotypes “epistacy.” During the ensuing decades, population geneticists have focused more and more on the potential importance of these interactions, all the while spawning their own terminology, which has frequently remained quite distinct from that of the rest of genetics (Wade 1992a; Phillips 1998). Phillips (1998) has recommended against using “epistasis” as a generic term to describe gene interactions. We generally try to follow this approach by using “epistasis” only when it is preceded by a modifier that describes the specific nature of the interaction (e.g., “synergistic epistasis”). Using the adjective “epistatic” is still appropriate because its meaning is usually clear from the context and there really is no practical alternative.

Population geneticists generally focus on the form of genetic interactions among mutant alleles that occur within a reference or average genetic background. If, for a particular phenotypic trait, there are no interactions among alleles, then one can imagine either that each allele adds a specific quantity to the trait or that each allele causes the trait to change by a specific fraction. In the first case, we say that an interaction is absent on an additive scale, while in the second case it is absent on a multiplicative scale. For any particular trait, it may not be obvious whether there is a natural scale to use in measuring gene interaction. With fitness, there is a theoretical reason to adopt a multiplicative scale: selection does not build up correlations among alleles at different loci when these correlations are initially absent if there is no interaction on a multiplicative scale (Felsenstein 1965).

Because gene interactions can take many forms, numerous terms have been used to describe them (Box 2.1). Most commonly in the evolutionary literature, epistasis is said to be synergistic (Crow 1970) or reinforcing (Kimura and Maruyama 1966; Crow and Kimura 1970, p. 80) if the effect of a mutation is stronger in the presence of other mutations than in their absence. Similarly, epistasis is defined as antagonistic or diminishing returns (Crow and Kimura 1970, p. 80) when the effect of a mutation is weaker in the presence of other mutations than in their absence. Box 2.1 shows that it is often easier to view gene interactions as causing either positive or negative deviations in a double mutant as compared to the product of the single-mutant fitnesses (Feldman et al. 1997). Unfortunately, the correspondence between the terms positive/negative and synergistic/antagonistic is not one-to-one but depends on whether beneficial or deleterious mutations are under consideration (Fig. 2.1 and Box 2.1). A further source of confusion is that the terms synergistic, antagonistic, and diminishing returns all refer to how mutations act in combination and do not refer to how they affect fitness. For example, although “synergy” might be thought to imply a beneficial effect on fitness, synergistic epistasis among deleterious mutations actually decreases fitness (Box 2.1). Again, there is some theoretical rationale for choosing among these sets of definitions. Positive linkage disequilibria develop among mutant alleles whenever epistasis is positive, whether or not the mutations are beneficial or deleterious (Eshel and Feldman 1970; Barton 1995a; Feldman et al. 1997). Similarly, negative disequilibria develop with negative epistasis. In contrast, the type of disequilibria that develop under synergistic or antagonistic epistasis depends on whether beneficial or deleterious mutations are

---

**Box 2.1 Classification of types of gene interactions in haploids**

Using the simplest two-locus model, we set the fitness of the AB haplotype at 1 and write the fitness of a haplotype with alleles a and b as \( W_{ab} = (1 + s_a)(1 + s_b) + \epsilon \), where \( s \) is the single-mutant effect and \( \epsilon \) is the epistatic deviation (since fitnesses must be nonnegative, parameters within the shaded region are impermissible; Fig. B2.1). When \( s \) is positive, then selection favors that allele when it is found with the wild-type allele at the other locus (A or B), while the opposite is true when \( s \) is negative. The interaction is defined as positive (\( \epsilon > 0 \)) or negative (\( \epsilon < 0 \)) epistasis, depending on whether the fitness of an individual with several mutations is higher or lower than the product of the individual fitnesses for each mutation considered separately (Feldman et al. 1997).

For deleterious alleles (\( s < 0 \); left side of Fig. B2.1), compensatory epistasis occurs when the fitness of the double mutant (ab) is greater than the fitness of the single mutants but not as high as the nonmutant (AB) (Kimura 1985; Phillips 1996; Stephan 1996). Supercompensatory epistasis occurs when the double mutant (ab) is also more fit than the nonmutant (AB); in this case, a peak shift would be required for the population to evolve a higher fitness (Wright 1977; Whitlock et al. 1995). Synergistic epistasis occurs when the double mutant is even less fit than predicted from the product of the single-mutant fitnesses, while diminishing returns epistasis occurs when the fitness of the double mutant is greater than this product, but still not as great as the single-mutant fitnesses.

For beneficial alleles (\( s > 0 \); right side of Fig. B2.1), decompensatory epistasis occurs when the fitness of the double mutant (ab) is less than the fitness of either single mutant; in this case, “beneficial” alleles a and b cannot both spread to fixation. Here, synergistic epistasis is characterized by the double mutants having even greater fitness than the product of the single-mutant fitnesses, while in diminishing returns epistasis the double mutant has lower fitness than this product, but still greater than the single-mutant fitnesses.

When one locus completely masks the effects of the other locus (\( s = 0 \); center of Fig. B2.1), synthetic deleterious loci are generated with negative epistasis (\( \epsilon < 0 \); Phillips 1998).
and Johnson 1998) and synthetic advantageous loci are found with positive epistasis ($\varepsilon > 0$; Slackin 1995) calls this complete epistasis].

A few numerical examples that assume symmetrical selection on both loci are presented in Table B2.1. Variance in epistatic effects matter because different pairs of loci may fall into different regions within this graph, thereby generating evolutionary behavior that can be qualitatively different from that predicted by the average level of epistasis.

<table>
<thead>
<tr>
<th>Type of Interaction</th>
<th>$\varepsilon$</th>
<th>$W_{ab}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deleterious mutations ($s = -0.2$; $W_{ab} = W_{aa} = 0.8$; $W_{ab}$ with no epistasis $= 0.64$):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Synergistic</td>
<td>-0.1</td>
<td>0.54</td>
</tr>
<tr>
<td>Diminishing returns</td>
<td>0.1</td>
<td>0.74</td>
</tr>
<tr>
<td>Compensatory</td>
<td>0.2</td>
<td>0.84</td>
</tr>
<tr>
<td>Supercompensatory</td>
<td>0.5</td>
<td>1.14</td>
</tr>
<tr>
<td>Advantageous mutations ($s = 0.2$; $W_{ab} = W_{aa} = 1.2$; $W_{ab}$ with no epistasis $= 1.44$):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decompensatory</td>
<td>-0.3</td>
<td>1.14</td>
</tr>
<tr>
<td>Diminishing returns</td>
<td>-0.1</td>
<td>1.34</td>
</tr>
<tr>
<td>Synergistic</td>
<td>0.1</td>
<td>1.54</td>
</tr>
<tr>
<td>Synthetic deleterious</td>
<td>-0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Synthetic advantageous</td>
<td>0.1</td>
<td>1.1</td>
</tr>
</tbody>
</table>

Figure 2.1 Change in fitness under an accumulation of mutations. A standard approach to investigating the influence of gene interactions on the evolution of sex and recombination is to see how fitness declines as the number of mutations within an individual increases. Here, epistasis is indicated by a departure from a multiplicative relationship among the mutant effects. The distinctions between the different views of gene interaction illustrated in Box 2.1 are also revealed here. Solid curves illustrate multiplicative interactions among mutants (top: beneficial; bottom: deleterious). Dashed and dotted curves represent different forms of epistasis.

Within the same individual will be simultaneously subject to these sometimes opposing forces.

**Measures of Gene Interaction**

With so many different terms being used to describe gene interactions, it is not surprising that there are a number of different approaches to the estimation of epistatic effects. There are many points of view about how this estimation should take place, which cannot be fully explored here (see Tachida and Cockerham 1989a; Phillips 1998; see Cheverud, chap. 4, this volume; Goodnight, chap. 8, this volume). For our purposes, the fundamental distinction is one between the epistatic effects that exist at the level of genotypic values and the epistatic variance found within populations. Epistatic effects emerge from the molecular and physiological context in which the genes are expressed. We abstractly describe these interactions using parameters such as the $\varepsilon$ given in Box 2.1. Epistatic variance is estimated as a fraction of the total amount of genotypic variation found in the population. Variance estimates are highly dependent on the genotypic frequencies found in the population and are frequently confounded by different sources of genetic effects (e.g., additive, dominance, and epistatic effects; Cheverud and Routman 1995). Goodnight (1995; see also chap. 8, this volume) describes how one can view changes in additive genetic variance in the presence of gene interactions as changes in the additive effects of alleles with changes in the relative frequency of the genetic
variable in epistatic effects discussed here is not dependent on allele frequency in the same fashion.

It is our view that epistatic variance is so removed from the actual interactions as to be basically useless for determining the impact of gene interactions on evolutionary processes. In particular, it is possible for there to be substantial amounts of gene interaction but for the epistatic variance to be near zero (e.g., Whitlock et al. 1995). This is partly because of statistical difficulties in the estimation of interaction effects (Wade 1992b) and partly because estimation of epistatic effects is based on taking the average over the entire genome. Most important, the average effects of alleles are defined in such a way that epistatic variance tends to be extremely small at equilibrium points, even if there are substantial genetic interactions underlying these effects (Whitlock et al. 1995). Since it is the effects themselves that are fundamental for our understanding of the evolutionary importance of gene interactions, any average over loci obscures the different evolutionary possibilities inherent in the interactions among any given set of loci. Moreover, variation in how gene interactions occur among loci in their interactive effects (even within an individual), whereas the epistatic variance describes the aggregate differences among individuals that cannot be accounted for using an additive model. Variance in epistatic effects dwells at the functional level of the interactions, while epistatic variance is a summary statistic far removed from the interactions themselves.

Genetic Interactions Are Predicted to Be Variable across Pairs of Loci

Metabolic control theory (Kacser and Burns 1981; Keightley 1989) predicts strong nonlinear interactions between mutant alleles at the same (dominance) or different (epistasis) loci. Szathmary (1993) has shown that both synergistic and antagonistic epistasis may result at the phenotypic level from additive interactions at the metabolic level. For example, if a phenotype (e.g., fitness) is proportional to flux through a pathway, a deleterious mutation that decreases the activity of one enzyme increases the importance (i.e., control coefficient) of that enzyme, which reduces the effect of deleterious mutations at other loci (diminishing returns epistasis). On the other hand, if there is an optimal flux through a pathway, synergistic epistasis may occur (but see G. P. Wagner et al. 1998). Consequently, metabolic control theory predicts variability in the ways that alleles interact, depending on the exact relationship between phenotype and the products of a pathway. [See Niederberger et al. (1992) for an empirical demonstration of these ideas.]

Developmental genetics also predicts variability in the ways that genes interact. The feedback mechanisms, gene regulation, and activation cascades inherent in development each create interactions among alleles whose form depends on the specifics of the developmental system. Indeed, the existence of extensive epistasis (in the classical sense) has provided a useful tool for ordering genes in development pathways (Avery and Wasserman 1992). Recent models that attempt to integrate developmental regulation with evolutionary change have predicted the emergence of gene interactions as a major feature of the evolution of developmental systems (Gibson 1996; G. P. Wagner et al. 1997; S. H. Rice 1998a, and chap. 5, this volume). Developmental systems are therefore expected to display not only gene interactions per se but also an extensive range of epistatic effects.

Genetic Interactions for Fitness Are Variable across Pairs of Loci

Mutation Analysis

Although numerous studies have found evidence for genetic interactions between genes or chromosomal regions (reviewed in Whitlock et al. 1995), relatively few studies have looked directly at the extent of variation in gene interactions between different sets of loci. Nevertheless, each of the studies that has looked at this has found strong evidence that the ways in which alleles interact is highly variable. In two elegant experiments with Escherichia coli, Elena and Lenski (1997) investigated the fitness consequences of new mutations by inserting transposable elements into random positions within the bacterial genome. In their first experiment, they measured fitness (growth rate) as a function of the number of insertion mutations in 225 clonal lineages. Fitness declined with an increasing number of mutations, but there was no evidence for gene interactions, either on an additive or multiplicative scale. There are, however, two potential explanations for this result: either gene interactions are rare or the interactions are highly variable and tend to cancel each other out. The second experiment by Elena and Lenski (1997) provided strong evidence in favor of the second explanation. In this experiment, they compared the fitnesses of wild-type cells, mutant cells that carry single transposable elements, and recombinant mutant cells that carry two of these elements. Out of 27 double-mutant lineages, 14 had a fitness that was significantly different than the product of the single-mutant fitnesses! The reason that no gene interactions were observed on average was that these significant epistatic effects tended to be positive as often as they were negative. This experiment highlights the fact that observation of little or no epistasis on average may indicate nothing about the prevalence and extent of genetic interactions.

Nearly identical conclusions can be drawn from an experiment by de Visser et al. (1997a) with the filamentous fungus Aspergillus niger. In this experiment, de Visser et al. took seven known marker mutations (two resistance and five auxotrophic mutations) and crossed them to obtain as many combinations of markers as possible. They succeeded in isolating and measuring the fitness of 73% of the possible multiple-marker strains. Plotting mean fitness on a multiplicative scale showed that mean fitness decreased linearly with the number of mutations, with no hint of epistasis on average. Nevertheless, genetic interactions between markers explained a significant portion of the variation in fitness between strains, thus suggesting that interactions were present but tended to cancel each other out. Finally, the authors compared double-mutant fitnesses to the product of the component single-mutant fitnesses for 14 cases where complete data was
available and where fitness declined with mutation number. Both antagonistic and synergistic interactions were evident among the mutations, although antagonistic interactions were more common. Although only one interaction was significant after a Bonferroni correction, the p-values for the interaction terms were less than .25 in all 14 cases, which is highly unlikely (< 10^{-8}) under the null hypothesis that pairs of mutations do not interact.

As a final example, Clark and Wang (1997) examined lines of Drosophila melanogaster into which P-elements had been inserted. Their focus was not on fitness but on 16 metabolic phenotypes (e.g., glycogen content and activity of fatty acid synthase). Single-P-element insertions against a uniform genetic background were used in a series of dihybrid crosses to generate all possible two-locus, two-allele genotypes. Interactions between eight pairs of P-element insertions were measured for each of the 16 traits. Out of the 128 tests, 35 were significant, with the form of interaction being highly variable (cf. Clark and Wang 1997, Figs. 2–5). This variability was so pronounced that the classification system of epistasis clearly breaks down—both positive and negative interactions were observed between the same two mutations, depending on whether the mutations were heterozygous or homozygous.

QTL Studies

In principle, the data emerging from recent studies that investigated the loci that underlie quantitative genetic variation (quantitative trait loci or QTL) should shed a great deal of light on the prevalence of gene interactions. For example, Long et al. (1995), in a study of bristle number variation in Drosophila melanogaster, found a large number of significant epistatic interactions among the observed QTL. The epistatic effects showed a great deal of variability from locus to locus, frequently changing sign even among comparisons that involved the same region. Similarly, Routman and Cheverud’s (1997) investigation of QTL that underlie body-weight differences in mice revealed extensive interactions among QTL regions that varied in a complex fashion from locus to locus. Again, particular loci often interacted differently with several other sets of loci, with the epistatic effects frequently changing sign from interaction to interaction. Effects also varied considerably depending on whether a given locus was heterozygous or homozygous. (The influence of gene interactions on quantitative traits is covered in much more detail in Cheverud, chap. 4, this volume; Templeton, chap. 3, this volume.)

A major difference between QTL analyses and the mutational approaches discussed in the preceding section is that the genetic background is not uniform in QTL analyses. Effects must be inferred by statistical averaging over all of the other alleles that segregate in the cross. Even in very large experiments, the tremendous number of possible genotypes means that many loci will be in linkage disequilibrium. This leaves current QTL estimates of epistatic effects closer to statistical averages than to the effects themselves. Nevertheless, these studies do highlight the very real existence of gene interactions and are consistent with these interactions being highly variable from locus to locus (and probably from allele to allele). The QTL studies will need to progress to the level of precision of the mutational approach before all of the potentially confounding factors can be sorted out (Phillips 1999).

Genetic Interactions Affect the Course of Evolution

Given the number of recent studies that highlight the existence of variability in epistatic effects, what are the evolutionary consequences of this variability? In the next sections, we consider the impact that variable epistatic interactions are likely to have on mutation load, evolution of recombination, drift of deleterious mutations, and speciation. Although our treatment is by no means exhaustive, we argue that each of these subjects can be strongly influenced by how genetic interactions are modeled.

Mutation Load

Although mutation is the ultimate source of the genetic variation necessary for evolutionary change, the bulk of mutations that affect fitness are deleterious (Simmons and Crow 1977). Natural selection opposes the continual production of deleterious mutations, but does so through a reduction in the average viability or fecundity of a population. The loss in fitness due to segregating mutations is known as the mutation load, after Muller (1950). The mutation load ($L$) can be defined as $1 - W$, where $W$ is the average fitness of a population in the presence of deleterious mutations relative to an average fitness of 1 for a hypothetical population free of deleterious mutations (see Crow 1970, 1993 for more details). The mutation load depends strongly on the genome-wide deleterious mutation rate ($U$). In the absence of fitness interactions among loci (i.e., when selection is multiplicative), the mutation load equals $1 - e^{-U}$ (Kimura and Maruyama 1966). In this case, the load becomes intolerably large when $U$ is greater than about 1. Gene interactions are known to have an important influence on the mutation load of a population (reviewed by Crow 1970, 1993), but the impact of variability in epistatic effects has thus far been largely ignored.

The mutation load has been shown to decrease in various models that incorporate synergistic epistasis, but each of these calculations assumes that interactions among loci are uniform in strength. For instance, with truncation selection, it is assumed that individuals with fewer than $k$ mutations have a fitness of 1 but that any individual with $k$ or more mutations dies; regardless of where in the genome these $k$ mutations occur (Kimura and Maruyama 1966; Kondrashov 1982). The reason for this is, of course, that keeping track of all possible genetic interactions among loci is difficult. In this section, we look at two simplified models that explicitly assume that gene interactions differ in strength between sets of loci to demonstrate that the mutation load is sensitive to assumptions made about epistatic interactions.

First, consider a haploid population in which genetic interactions occur only within distinct pairs of loci. That is, if loci $A$ and $B$ form an interacting pair, then the fitness of an allele at locus $A$ depends on the genotype at locus $B$ but is independent of the alleles carried at all other loci. Imagine the fitness contributed
by the \( j \)th pair of loci is equal to 1 (if there are no mutations in the pair), \( 1 + s_j \) (with a single mutation for the pair), or \((1 + s_j)^2 + \varepsilon_j \) (if both loci carry mutant alleles). The fitness of an individual is then calculated as the product of the fitnesses determined for each pair of loci. For deleterious mutations and synergistic epistasis, both \( r \) and \( \varepsilon \) are negative (Box 2.1).

Assuming a per locus mutation rate of \( \mu \), \( \omega \) pairs of interacting loci, a genome-wide mutation rate of \( U \), and scaling epistasis by the strength of selection \( [\varepsilon^* = \varepsilon_j/(\gamma_1 \gamma_2)] \), we obtain the following approximation for the mean fitness in the population:

\[
\ln(W) \approx -U - \omega \mu^2 \varepsilon^* - 2 \omega \mu^2 \varepsilon^* (\text{Var}[\varepsilon^*] + \varepsilon^*^2)
\]

where an overbar denotes the average value for a quantity (see the appendix).

As can be seen from the second term, synergistic epistasis \((\varepsilon^* < 0)\) decreases the mutation load of a population (i.e., the second term is positive). But, as can be seen from the third term, for a given average level of epistatic interaction \((\varepsilon^*)\), variability in the extent of epistasis \(\text{Var}[\varepsilon^*]\) decreases mean fitness and increases the mutation load experienced by a population (i.e., the third term will always be negative, no matter what the average form of epistasis).

This conclusion rests upon a Taylor series approximation that assumes weak epistatic interactions. With extensive variability in epistatic effects, some pairs of loci may experience very strong synergistic epistasis (with a lessened load), and some may experience such strong diminishing returns epistasis that multiple mutations become compensatory or even beneficial (Box 2.1). If a population can traverse the fitness valley created by mutations that are deleterious on their own but beneficial in combination (Wright 1977; Michalakis and Slatkin 1996; Phillips 1996), then a portion of the mutation load may be converted into adaptive evolutionary change. In this case, variation in epistatic effects could potentially allow an increase in the mean fitness of a population.

The second approach we take is to fit different models of gene interaction to data and then to measure the sensitivity of mutation load estimates to these different models. The classic empirical example for epistasis is Mukai's (1969) mutation accumulation experiment with Drosophila melanogaster. Mukai allowed mutations to accumulate in several heterozygous lines by severely restricting the population size of each line. Periodically, the second chromosome was extracted from each line and made homozygous. The average viability was measured over time for 72 such lines (Fig. 2.2). Viability declined over time at an increasing rate, suggesting synergistic epistasis. A quadratic equation, \( W = 1 - ax - bx^2 \), for fitness as a function of the number of mutations \((x)\) provides an excellent fit to the data (Mukai 1969), as does a quadratic function for the natural log of fitness, \(\ln(W) = -ax - bx^2\) (Crow 1970). However, many other models of gene interaction can also fit the data. As an extreme example, consider a model of genetically interacting clusters of loci with truncation selection. (By the term “cluster” we do not mean to imply anything about the physical relationship of these loci, but rather that they interact with other members of the same set, or cluster.) Imagine \( C \) clusters of interacting genes within the genome. Within a cluster, if there are fewer than \( k \) mutations, fitness remains 1, but it falls to zero with \( k \) or more mutations. There are no fitness interactions among clusters. Hence, the fitness of an individual equals 1 only if the fitness due to each cluster equals 1. Surprisingly, we can get an equally good fit to the data simply by varying the values of \( C \) and \( k \) (Fig. 2.2).

The mutation load of a population, however, depends strongly on which of the many models that fit the data actually describes the underlying genetic interactions. The models (see Fig. 2.2) give different estimates for the mutation load (Table 2.1), even though they all provide good fits to the observed decline in fitness in a mutation accumulation experiment. Certainly, mutational loads estimated from genetic data will depend on the assumptions made about the distribution of epistatic effects. This observation highlights an important point: just because a simple function that treats all mutations as equivalent can fit data, it matters what \( C \) and \( k \) are.

![Figure 2.2](https://example.com/fig2.2.png)

**Figure 2.2** Fitness as a function of the number of homozygous mutations carried on the second chromosome of Drosophila melanogaster. Data (•) are from Mukai's (1969) mutation accumulation experiment. Assuming a mutation rate for the second chromosome of 0.1 (Crow 1970), the curves describe the loss in fitness as a function of the number of mutations accumulated under different models: multiplicative selection (thick solid curve), quadratic selection (thin solid curve), quadratic selection on a log scale (dotted curve), and the clusters model with truncation selection (dashed curve). The parameters in the three epistatic models were chosen to best fit the data (see Table 2.1) and provide equally good fits to the data.

**Table 2.1 Estimated mutation load under different models.**

<table>
<thead>
<tr>
<th>Model of Selection</th>
<th>Best Fit</th>
<th>( U = 0.1 )</th>
<th>( U = 0.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiplicative</td>
<td>0.0952</td>
<td>0.393</td>
<td></td>
</tr>
<tr>
<td>Quadratic</td>
<td>0.0669</td>
<td>0.236</td>
<td></td>
</tr>
<tr>
<td>Log quadratic</td>
<td>0.0638</td>
<td>0.260</td>
<td></td>
</tr>
<tr>
<td>Clusters</td>
<td>0.0498</td>
<td>0.230</td>
<td></td>
</tr>
</tbody>
</table>

*For the models of gene interaction considered, the parameters were varied until the best fit to Mukai's (1969) data was obtained (see Fig. 2.2). Since the data are based on fitness effects of homozygous chromosomes, only the mutation load at equilibrium in a sexual, homozygous diploid population is considered. Numerical entries give the mutation load predicted by the models at different genome-wide mutation rates, \( U \).*
this does not necessarily mean that the function is a good description of the ways in which genes interact. (Indeed, there is some question as to whether it is valid to treat the mutation rate as constant; Keightley 1996; Nuzhdin et al. 1997.) The cluster model is clearly extreme, but so too is a model that ignores the fact that some genes interact more strongly than other genes. W. R. Rice (1998) advocates circumventing these issues by directly assaying epistatic relationships in closely interacting pathways.

Evolution of Recombination

How does the recombination rate among loci enhance or mitigate the influence of genetic interactions, and how might this influence affect the evolution of recombination itself? Theories about the evolution of sex and recombination have long focused on the important role played by gene interactions (reviewed by Feldman et al. 1997; Barton and Charlesworth 1998; Otto and Michalakis 1998; see also Peters and Lively, chap. 6, this volume). Although relatively unexplored, variation in epistatic effects has an important impact on the outcome of these theoretical models.

Sex and recombination destroy genetic associations, thereby generating rare, and potentially favorable, genotypes. Yet, sex risks the dismantling of genotypes that survived selection to produce less-fit recombinant offspring. The balance between these two factors determines when sex and recombination will evolve within a population (Barton 1995a). With negative epistasis among deleterious mutations, selection reduces the genetic variation within a population, because individuals that carry many mutations have an especially low fitness. Under these circumstances, sex tends to regenerate extreme genotypes by recombining intermediate genotypes. This leads to a higher representation of the "best" genotype (with few deleterious mutations) among those individuals that have sex, and, therefore, the evolution of sex can be favored by negative epistasis. However, sex also generates more of the "worst" genotypes, and the average fitness of recombinant offspring is often lower than that of nonrecombinant offspring, especially when epistasis is strong. This is important, because it works against the evolution of sex as negative epistasis becomes stronger (see Fig. 2.3).

With positive epistasis, selection is less effective at eliminating individuals with a high number of mutations, and more genetic variation is maintained within a population. Sex now tends to recombine extreme genotypes to produce offspring with an intermediate number of mutations. This has a double disadvantage: genotypes that undergo sex more often become disassociated with the "best" genotype and tend to produce offspring that are less fit on average because of the breakup of the associations favored by selection. These disadvantages worsen as epistasis becomes positive.

Since the advantages of sex diminish as negative epistasis becomes stronger, while the disadvantages of sex intensify as positive epistasis becomes stronger, variance in the amount of interaction between pairs of loci always reduces the amount of selection that favors increased sex within a population (Otto and Feldman 1997). Very similar results can be obtained for the evolution of recombination among epistatically interacting beneficial alleles (Barton 1995a).

![Figure 2.3](image-url) The influence of negative epistasis on the evolution of recombination in a three-locus model. The model and parameter space is located toward the very center of Fig. B2.1. Shaded areas show regions in which higher recombination rates would be expected to evolve based on the interaction between two loci (A and B) subject to viability selection, where a third locus (M) modifies recombination rates. The recombination rate is initially assumed to be 0.1 between loci A and B and between loci B and M. Even if, on average, epistasis falls within the shaded region, variance in epistatic effects among pairs of selected loci (causing variation along the y-axis) will cause some pairs of loci to fall outside of the shaded region. This will weaken or negate selection for increased recombination (Otto and Feldman 1997). Curves are based on eq. 18a from Barton (1995a) for deleterious alleles held at mutation-selection balance (s < 0; left side) and eq. 17 from Barton (1995a) for beneficial alleles that spread through a population (s > 0; right side).

Therefore, for sex and recombination to evolve as a result of gene interactions under directional selection, negative epistasis with little variance in the amount of epistasis is required (Fig. 2.3).

Drift and Fixation of Deleterious Mutations

Mutations that arise in a population may drift to fixation even if they are deleterious, especially if the effective population size is small. This process has been termed "drift load." This happens most frequently if the selective effect of the mutation (s) is small and the effective population size (N) is also small, particularly if 4Ns is less than 1 (Wright 1931; Crow and Kimura 1970). If these mutations continue to accumulate, the mean fitness of the population can be reduced substantially. It has been proposed that this reduction in mean fitness can lead to further decreases in population size, which, in turn, makes it more likely that further deleterious mutations can fix, and so on as the population slowly collapses upon itself (Kondrashov 1995; Lande 1995; Lynch et al. 1995a, 1995b). This "mutational meltdown" occurs much faster in an asexual than in sexual species and, therefore, has been proposed as a major reason for the advantage of sexual over
asexual lineages (see Peters and Lively, chap. 6, this volume). Furthermore, avoidance of the mutational meltdown has been suggested as an important consideration in the preservation of biodiversity (and has, in fact, been a major consideration in management recommendations on the subject), as it would accelerate the loss of species with reduced population sizes (Lande 1995; Lynch 1996).

Gene interactions affect the rate of accumulation of deleterious alleles, but in different ways in asexual and sexual populations. In asexual populations, fixation of a deleterious mutation within a lineage irrevocably reduces the fitness of that lineage because of the lack of recombination (Muller 1964). This process, known as Muller’s ratchet, will proceed slowly and inexorably until even the fittest genotype in the population has accumulated many deleterious alleles. The smaller the effect of these alleles, the easier it is for the ratchet to click, but the less it matters. Therefore, alleles of moderately small effect can be the most dangerous for the health of an asexual population. With synergistic epistasis, the incorporation of more and more mutations within a population causes new mutations to have larger effects and, consequently, to be less likely to become established. As a result, epistasis of this sort can slow or halt the progression of Muller’s ratchet in asexual populations (D. Charlesworth et al. 1993; Kondrashov 1994). However, Butcher (1995) has shown that this slowdown in the ratchet is dependent on the critical assumption that the magnitudes of selection and epistatic interactions are constant. If there is variation in these quantities, then there will always be alleles whose overall fitness effects are small enough to avoid selective elimination, and the ratchet will continue. Variance in epistatic (and additive) effects causes Muller’s ratchet to progress differently than would be expected by the average of the effects alone.

With sexual species, fixation of deleterious alleles by drift can cause mutational meltdown as well (Kondrashov 1995; Lynch et al. 1995a, 1995b). Kondrashov has shown that strong synergistic epistasis can substantially reduce drift load when the product of the effective population size and the per nucleotide mutation rate is less than 1. Similarly, Schultz and Lynch (1997), using a different fitness function and distribution of allelic effects, have shown that average synergistic epistasis can reduce the probability of extinction due to the accumulation of deleterious alleles, although by less than an order of magnitude.

The effects of variance in epistatic effects on drift load are complex and have not been well studied. It is known that variance in the average effect of mutations can substantially change the time to extinction of a population (Schultz and Lynch 1997). For the following reasons, it also seems clear that the time to extinction should be very sensitive to the distribution of epistatic interactions. Following the fixation of some deleterious mutations, the drift load will be reduced if new mutations interact synergistically with existing mutants (because the new mutations are less likely to fix). The drift load will be increased if the old and new mutations interact antagonistically (because the new mutations are more likely to fix). Finally, it will be reduced again if the subsequent mutations compensate for the mutations that have come before (because the new mutations are then actually beneficial). A mixture of these classes of mutations could very well have a beneficial effect on average, thereby reducing or perhaps even alleviating the problems of drift load. Yet, this would not necessarily be predicted based on the average degree of epistasis. For example, if the average epistatic interaction is antagonistic, then, in the absence of variation in epistatic effects, the problems of drift load would be exacerbated. On the other hand, if variation around this average is substantial, and both synergistic and compensatory mutations are common, the opposite result might be the case. Clearly, the distribution of epistatic effects matters, not merely the mean of their values.

Multiple Peaks, Multiple Evolutionary Trajectories
One of the most striking facts about biological organization is that individuals are organized into more or less independent species, and are capable of mating with members of their same species but are incapable of producing fit offspring by mating with members of most other species. The strong implication of this is that there is a great deal of gene interaction involved in the determination of fitness, particularly when allele combinations are not constrained by selection to work well together, as is the case when alleles arise in different populations or species (Whitlock et al. 1995). How these genetic differences come about is still a topic of great debate (Coyne 1992; see Johnson, chap. 12, this volume). In order for reproductive isolation to evolve, it is essential that two daughter populations take different evolutionary trajectories after they begin to become isolated. This means that there must be alleles that interact strongly enough to produce speciation; these alleles either segregate within the ancestral population or arise subsequently due to mutation.

If we look at the average interaction effects measured within populations, however, the effects that we see are often at or near zero (see above). Variance among allele pairs in the amount of epistatic interaction is required for the evolution of reproductive isolation. Speciation could not occur if alleles always interacted in the weak fashion predicted by the average (i.e., some fraction of the alleles must fall into the upper left quadrant of Fig. B2.1, even if the average $e$ is closer to the central axis). As a result, in order to understand the genetic basis of speciation and the population genetic processes that lead to speciation, we must understand not only the average interaction of pairs of novel mutations, but also the distribution of these interaction effects. If deleterious interactions are rare among alleles that are capable of working well in some other genetic context, then speciation will proceed slowly. If these mutations are common, then this would accelerate the pace of speciation (Orr 1995; Wade and Goodnight 1998).

The adaptive landscapes that populations traverse as they diverge are complex, multidimensional entities (Whitlock et al. 1995). Variation in epistatic effects creates peaks and fissures of different heights in different regions of the landscape. However, populations need not necessarily traverse adaptive valleys to diverge from one another (Dozhansky 1936; Muller 1939; Orr 1995; Whitlock et al. 1995). The nature of the interaction can be so complex that the epistasis between any two loci depends on variation at other loci (e.g., Cabot et al. 1994). In this case, the total set of interactions can generate a "holey" landscape on which species are separated by paths around the fitness valleys (Gavrilets and Gravner 1997; Gavrilets et al. 1998). More fundamentally, if gene interactions are pervasive and highly variable, then the evolutionary context of each new
mutation is dependent on the suite of mutations that have been fixed beforehand. The mutational landscape (Gillespie 1984) becomes more central as gene interactions become more common. Waiting for the proper set of interacting mutations to enter a population can have a dramatic effect on the rate of evolutionary change (Gillespie 1984; Phillips 1996). The influence of variation in epistatic effects across the mutational landscape has yet to be fully considered.

Discussion

To understand the importance of variability in the effects of interactions between multiple loci, consider first the influence of variability in the effects of single loci. For adaptive evolution, alleles with large positive effects on fitness should be fixed in a population at a faster rate than alleles with smaller effects, although mutations with larger effects may be rarer overall (Fisher 1930; Kimura 1983; Orr 1998). Similarly, in the elimination of deleterious alleles, those alleles with large negative effects on fitness should be rapidly eliminated from a population, while mildly deleterious alleles can segregate for a much longer time, perhaps even reaching fixation due to drift (Ohta 1973; Kimura 1983). It is self-evident that the rate of evolution depends on the mutational distribution of allelic effects, which, on fundamental grounds, we would expect to show a fair amount of variability.

For the case of genetic interactions, the influence of variability is much more subtle. If some interactions enhance the effects of some loci while other interactions are masking, the outcome of selection will depend on the relative frequency of each of these types of interactions. The genotype of the individual establishes the context in which the effects of any allele are expressed. From the point of view of an allele, this context is continually changing after each bout of recombination. Since different sets of loci can potentially generate very different types of gene interactions, variability in epistatic effects across the genome can strongly influence the average attributes of both individuals and alleles, especially as the population context of the allele also changes (Wade 1992b; Goodnight 1995, and chap. 8, this volume). The essential point is that many models that incorporate the influence of gene interactions have different outcomes according to the nature of the interaction (e.g., synergism vs. antagonism). Variance in effects across loci means that different sets of loci will potentially experience these separate domains simultaneously within the same individual. This has two fundamental consequences. First, looking at the average level of epistasis across an entire chromosome or within an individual may falsely suggest that there are no gene interactions because the different forms of epistasis balance each other out (e.g., Elena and Lenski 1997). Second, the decrease in apparent epistasis generated by averaging does not mean that gene interactions have no effect, because variance in epistasis itself can affect evolutionary outcomes (e.g., Fig. 2.3; Otto and Feldman 1997).

The design of studies to look for variable epistatic effects is complicated by the fact that the extremely high levels of genetic variation found in most populations make it difficult to separate out a specific set of interactions from the background of thousands of other potential interactions. As early as 1974, Lewontin (1974, p. 42) recognized this problem when he wrote with regards to isolating the effects of single loci, "there is simply no way to make a large number of individuals identically homozygous or heterozygous at one locus while keeping the rest of the genome segregating at random." The tremendous number of potential genetic backgrounds overwhelms the detection of the interaction between any particular locus pair. This churning genetic background helps to illustrate why Fisher focused so strongly on the average effects of an allele—the importance of other effects is likely to be lost in the sea of potential interactions. Indeed, the average effect of an allele will tend to dominate evolutionary change at any particular locus. Nevertheless, averaging over all possible interactions leads to the loss of important information. Variability in epistatic effects will matter to any phenomenon that depends on gene interactions in a nonlinear fashion, which is true for all of the examples that we have explored thus far.

The studies that have found variance in epistatic effects have done so by isolating genetic changes against a fixed background (Clark and Wang 1997; de Visser et al. 1997a; Elena and Lenski 1997). This was accomplished by utilizing targeted mutations to pinpoint the effects of specific loci. What is gained in this approach is precision in the estimate of epistatic effects for that particular allelic combination. What is lost is the overall population context in which these interactions must exist. The clear way out of this quandary is to test more and more loci against more and more genetic backgrounds. This is a potentially daunting task, but one that is becoming technologically feasible and that will get us closer to solving Lewontin's dilemma. We have yet to fully develop a conceptual basis for describing gene interactions that combines both the functional nature of genetic interactions and the population context in which those interactions exist (Phillips 1998; see Templeton, chap. 3, this volume). Understanding the nature of the distribution of epistatic effects will obviously be key to linking these two levels of analysis.

As evolutionary genetics continues to incorporate more functional genetics into investigations of epistasis, studies that dissect particular gene interactions will become more feasible. Inherent in this point of view is an explicit rejection of Fisher's (1930) hope for a set of rules, similar to the "ideal gas law," with which one could completely describe evolution by the use of summary parameters (e.g., the additive genetic variance). Many current models in evolutionary genetics, including many from quantitative genetics (e.g., Barton and Turelli 1989, 1991), demand more information regarding the particular nature of allelic effects, especially as they pertain to new mutations. It may be that after going through a period of atomizing these effects, we will turn again to summaries and averages. However, it seems evident that we need more knowledge about the nature of gene interactions and how these interactions vary across the genome before we can take this step, and we therefore need to focus our attention, at least for a while, at looking beyond the average.

Acknowledgments. We thank the editors for inviting us to participate and for their help with the chapter. We also thank Michael Lynch, Norman Johnson, and an anonymous reviewer for comments on the manuscript. This work was supported by grants from the
Appendix

The importance of variation in epistatic effects can be illustrated using a simple model with epistatic interactions. Assume that there are $\omega$ pairs of interacting loci and that the genotype of an individual is given by $G$, where $G_i$ is equal to 1 if locus $i = \{A \text{ or } B\}$ in the $j$th pair of loci ($j = \{1, \ldots, \omega\}$) carries a mutation and is zero otherwise. Let $s_{ij}$ measure the strength of selection at locus $i$ in the $j$th pair in the absence of another mutation within the pair, and let $e_{ij}$ measure the strength of the gene interaction when both loci in pair $j$ carry a mutation. Selection is multiplicative between sets of loci. The fitness of an individual is then given by

$$W_G = \prod_{j=1}^{\omega} [(1 + s_{A_i} G_{A_j})(1 + s_{B_i} G_{B_j}) + e_{ij} G_{A_i} G_{B_j}] \quad (2.1)$$

The mean fitness of a population can be calculated from the mean fitness based on each pair of loci. Assuming random mating, and following chromosomes through selection, recombination, and mutation, we obtain the mean fitness of the population as

$$\bar{W} = \prod_{j=1}^{\omega} [(1 + s_{A_i} p_{A_i})(1 + s_{B_i} p_{B_i}) + s_{A_i} s_{B_i} D_j + e_{ij} (p_{A_i} p_{B_i} + D_j)] \quad (2.2)$$

where $p_{ij}$ is the frequency of mutant alleles at locus $ij$ and $D_j$ is the genetic linkage disequilibrium between loci $A$ and $B$ in pair $j$. Implicit expressions for $p_{ij}$ and $D_j$ in terms of $\mu_{ij}$ (the mutation rate at locus $ij$) and $r_j$ (the recombination rate between loci $A$ and $B$ in pair $j$) can be obtained from the two-locus haploid selection model. These expressions can be used in a Taylor series approximation of the log of mean fitness. Keeping linear and second-order terms in $e_{ij}$ we get

$$\ln(\bar{W}) \approx \sum_{j=1}^{\omega} \left[ \ln(1 - \mu_{A_i}) + \ln(1 - \mu_{B_i}) - e_{ij} \frac{r_j \mu_{A_i} \mu_{B_j}}{s_{A_i} s_{B_i} \Psi} - e_{ij}^2 \frac{2(1 - r_j) r_j \mu_{A_i} \mu_{B_j}}{s_{A_i} s_{B_i} \Psi^2} \right] \quad (2.3)$$

where $\Psi = 1 - (1 - r_j)(1 + s_{A_i})(1 + s_{B_i})$. For ease of presentation, we have included only the leading order terms in the mutation rate. (Details of the analysis are available at http://www.zoology.ubc.ca/~otto.) To simplify eq. 2.3 further, we scale epistasis by the strength of selection, defining $e'_{ij} = e_{ij} / (s_{A_i} s_{B_i})$, and we assume that covariances among $e'_{ij}$, $s_{ij}$, $r_j$, and $\mu_j$ are small. Then, for weak selection, small mutation rates per locus, and a high average recombination rate ($r \approx 1/2; \Psi \approx 1/2$), we have the approximate result given in eq. 2.1.